基于多模态视觉语言模型的老旧社区环境特征解析*

——以重庆市中心城区为例

李彦锦[1] 罗 丹[1] 肖 竞[1] 李 玮[2]

（1. 重庆大学建筑城规学院，重庆 400045；2. 长春中海地产有限公司，长春 130000）

摘要：【目的】针对老旧社区环境特征识别中传统方法效率低、受主观因素影响较大的局限，提出一种融合多模态视觉语言模型（VLM）与带地理信息的众源图像的创新方法框架，识别与提取老旧社区环境特征，并对其进行量化与归类。【方法】以重庆市中心城区为研究区，通过构建众源图像数据集，利用多模态视觉语言模型提取图像语义转化为文本，结合 BERTopic 主题建模技术提取聚类，将聚类结果映射到地理空间分析特征分布和共现结果。【结果】提取 50 个聚类得到涵盖空间骨架、微观细节、绿化形式及场所氛围的 17 组环境感知特征主题，识别出 7 类典型的特征空间共现模式，为空间定向改造提供建议。【结论】老旧社区环境具有复杂的异质性特征，并受历史风貌与公共生活的深度影响；多模态分析框架能够有效对老旧社区环境图像进行语义理解，实现低成本、高通量且精细化的环境特征挖掘。未来可以进一步结合社会学数据构建完整的更新闭环，真正服务于“好房子、好小区、好社区、好城区”的系统性建设。

关键词：多模态视觉语言模型；老旧社区；众源图像；空间特征；重庆市中心城区

DOI：10.12169/zgcsly.2026.01.01.0001

Decoding Environmental Characteristics of Old Urban Residential Compounds Via Multi-modal Vision-Language Models: A Case Study of the City Core of Chongqing

Li Yanjing[1] Luo Dan[1] Xiao Jing[1] Li Wei[2]

(1. Faculty of Architecture and Urban Planning, Chongqing University, Chongqing 400045, China; 2. Changchun Zhonghai Property Co., Ltd., Changchun 130000, China)

Abstract: 【Objective】 To address the limitations of traditional methods in identifying environmental characteristics of old urban residential compounds, including low efficiency and susceptibility to subjective factors, this study proposes an innovative methodological framework integrating Multi-modal Vision-Language Models (VLM) with Geotagged Crowdsourced Imagery to have the environmental characteristics of old urban residential compounds decoded and interpreted, which is then quantified and classified. 【Method】 With the city core of Chongqing taken as the empirical study area, a dataset of crowdsourced images is constructed. VLMs are utilized to transcode visual semantics into textual descriptions, which are then processed using BERTopic modeling for clustering analysis. The resulting clusters are mapped onto geographic space to analyze the spatial distribution and co-occurrence patterns of environmental characteristics. 【Result】 The extraction of 50 clusters yields 17 thematic groups of environmental perception features, covering spatial skeletons, micro-scale details, greenery forms, and place ambiance. Furthermore, 7 typical spatial co-occurrence patterns of these characteristics are identified, providing

收稿日期：2026-01-01

* 基金项目：国家自然科学基金重点项目（52238003）；重庆市社会科学青年项目（2021NDQN64）

第一作者：李彦锦，男，2001 年生，硕士生，研究方向为风景园林规划与设计。E-mail：202415131221@stu.cqu.edu.cn

通信作者：罗丹，男，1989 年生，博士，副教授，研究方向为风景园林规划与设计。E-mail：luodan@cqu.edu.cn

suggestions for spatially targeted renovation. 【Conclusion】 The environments of old urban residential compounds exhibit complex heterogeneity, profoundly influenced by historical features and public life. The multi-modal analysis framework effectively enables semantic understanding of environmental images, achieving low-cost, high-throughput, and fine-grained feature mining. Future research can further integrate sociological data to build a complete loop of renewal, truly serving the systematic construction of "good houses, good neighborhoods, good communities, and good urban districts".

Keywords: multimodal vision-language model; old urban residential compound; crowd-sourced imagery; spatial feature; city core of Chongqing

当前，中国城镇化进程已从大规模增量建设转向以提质增效为核心的存量更新阶段[1]。老旧社区作为城市有机体的重要组成部分，其更新改造不仅是改善民生的关键举措，亦是提升城市治理现代化水平的核心议题[2]。2020 年，国务院办公厅印发《关于全面推进城镇老旧小区改造工作的指导意见》，明确提出应大力推进城镇老旧小区改造，其中社区环境改造工作主要聚焦于保障基础安全，筑牢宜居底线；完善环境与设施，提升生活便利性；推动智慧化升级，拓展社区服务功能；最终目标是切实提升居民的获得感、幸福感与安全感[3]。

老旧社区主要指我国城市中于 2000 年以前（尤其是 20 世纪 90 年代前）集中建成并投入使用的居住区[4]，其环境认知与管理长期面临严峻挑战；其空间形态往往复杂、破碎且处于动态变化之中；功能组成多元，评价标准因人群而异[5]；物质界面常与既有建成环境相互交织，居民自发改造利用的空间往往是地方活力的重要载体。传统调研方法如问卷、访谈与人工踏勘，在此类空间中具有明显局限：一方面效率不佳，难以实现大规模系统性普查[6]；另一方面主观性强，高度依赖个人经验与判断。因此，在更新实践中缺乏针对性的老旧社区空间识别路径，使得老旧社区的潜在价值未能被充分识别与利用，更新治理也难以做到精准施策。

深度学习与计算机视觉技术为城市空间感知提供了新路径。多模态视觉语言模型能够融合图像、文本与地理信息等多源数据，模拟人类对环境的综合理解对空间进行识别与解析。同时，实时、具身的众源地理数据，尤其是海量、细粒度的众源图像为自下而上刻画非正式空间提供了丰富的数据基础[7]，其中利用街景图像、社交媒体数据可以有效识别城市热点[8-9]，但如何系统融合多模态视觉语言模型与众源图像，构建从识别、特征量化到潜力评价的完整技术框架，并应用于老旧社区非正式空间这一特殊对象，仍是亟待探索的问题。

本研究将老旧社区环境定义为承载居民日常外部活动的空间场所，包含建筑实体以外的道路、广场、绿地、公共设施等功能空间及其构成要素，旨在提出并验证一个融合多模态视觉语言模型与众源图像数据的创新方法论框架，并以重庆市中心城区为典型区域，解析两个关键问题：1）如何基于多模态视觉语言模型从众源图像中有效识别与提取老旧社区环境特征？2）老旧社区环境特征应如何量化与归类，以深化其空间认知？

1 研究区概况

重庆中心城区作为典型的山地超大城市，其老旧社区在空间上呈现出“立体错落、高密度混合、新旧交织”的鲜明特征。这些社区多建于 20 世纪 80—90 年代或更早，普遍面临设施老化、公共空间不足等问题。受复杂地形制约，社区往往依山就势、布局紧凑且破碎化程度高，为复杂建成环境研究提供了极具挑战性与极具价值的样本。

作为国家首批城市更新试点城市，重庆正全面推进老旧小区改造。“十四五”期间，重庆共完成改造老旧小区 8 074 个，主要集中于中心城区，其环境品质提升是核心任务之一。此外，中心城区人口密集，经前期数据收集发现，海量众源地理数据（如街景图像、社交媒体照片、POI 兴趣点）广泛分布于老旧社区，为运用多模态视觉语言模型实现大范围、细粒度的环境特征挖掘提供了丰富数据基础。

2 研究方法

2.1 多模态模型挖掘老旧社区特征的潜力

通过挖掘海量的网络资料，研究者能够有效揭

示城市感知特征。现有研究主要基于图像或文本单模态聚类分析：文本数据上，传统方法多基于评论或网络游记，通过词频分析、语义网络或情感分析等技术提取关注热点及满意度评价[10-12]，但难以提取环境微观细节且缺乏精确地理坐标，无法追溯文本描述对象；图像数据上，基于深度学习模型提取图像的视觉特征向量，结合聚类算法识别出城市街道的视觉类型和美学质量分布[13-14]，但需要对抽象的图像聚类结果进行人为解释。为了突破单模态的局限，近期研究利用多模态模型将图像转换为文本描述，进而得到包含语义的聚类结果[15-16]，但现有研究多限定在特定历史街区或宏观城市层面的特征描述，尚未充分将多模态模型解析出的细粒度感知特征系统地映射到社区尺度的地理空间中，以指导微观层面的更新与治理。基于以上进展，本研究提出基于多模态模型的图文融合分析方法，利用大语言视觉模型对老旧社区 VGI 图像进行图像-文本转换，结合 Bertopic 文本主题建模和多模态方法提取特征并将其映射到地理空间中分析老旧社区的视觉空间特征。

2.2　数据收集与处理

2.2.1　构建数据集

六只脚是一个基于位置服务的户外平台，通过用户上传和下载 GPS 轨迹和带地理信息图像等方式提供信息资源和服务，包含大量人视角步行场景图像。本研究按老旧社区名称或街道名称等关键词，从六只脚平台收集 28 个老旧社区共计 2 995张含地理信息的照片，筛除图像质量较差、重复、非人视角和非外部环境的图像后，最终使用包含 1 588 张图像的数据集进行图像转文本和聚类分析。

2.2.2　利用多模态视觉语言模型图像转文本

以 GPT4o 模型为代表的多模态视觉语言模型被认为具有良好的视觉和空间理解能力[17]。考虑到选区样地的文化背景，采用 Doubao-Seed-1.6-vsion 250815 模型开展图像转文本工程，该模型在大规模的中文语料库上进行预训练，能够理解和生成自然、流畅的中文文本，以匹配老旧社区使用人群的文化语境；同时该模型在多个公开测试集中表现出前沿的多模态能力，能够准确识别和分析图像要素。通过分段构建提示词的方式让 VLM 模型按空间、细节、绿化、氛围 4 个维度进行文本描述：1）空间：空间核心骨架与界面，2）细节：地面肌理与微观物体；3）绿化：植物形态与绿化形式，4）氛围：生活痕迹与场所体验（表 1）。

表 1　提示词工程

类型	内容
前置信息	#角色 你是一位拥有敏锐观察力的“城市微更新观察家”和“空间叙事者”，任务目标是捕捉老旧小区图像中最独特、最具辨识度的视觉特征；描述将被用于语义聚类，请强调不同场景之间的“差异性”
风格约束	#风格 以描述“视觉质感”与“具体名词”为主，请勿使用“报告式”语言 ＊错误示例：“地面铺装平整，绿化较好” ＊正确示例：“暗红色的六边形透水砖路面上斑驳地散落着落叶，两侧不仅有高大的水杉，还杂乱地生长着野生的蕨类植物”
描述维度	#描述维度（用 === 隔开） 1）空间核心骨架与界面：请定义场所的空间属性，用连贯的笔触描述描述地面延伸方向和高低变化、空间围合的方向及围合界面的颜色、材料、特点 === 2）地面肌理与微观物体：具体描述地面的材料和状态，如“破碎且长满青苔的混凝土”；请列举场景中具体物体，不要概括，如“生锈的铁质晾衣杆” === 3）植物形态与绿化形式：禁止估算覆盖率，请描述植物的姿态，如“修剪整齐的灌木球”；描述绿化形式，如“砖块参差的规则种植池”“随意摆放的陶瓷花盆”亦或者其他组合形式 === 4）生活痕迹与场所体验 描述这里的氛围体验，避免评价功能，而是根据要素和使用痕迹来推断场所的氛围；将观察到的痕迹转化为“体验性的描述”，如“萧条废弃
负面约束	#负面约束 1）禁止使用评价性词汇：如“美观度一般”“包容性不足”“安全性良好”“绿化率 30%”等 2）禁止否定句：避免“无无障碍设施”，只描述存在的东西 3）禁止通用句式：避免“图片展示…”“可以看到…”等开头，直接开始描述
格式要求	#输出结构 严格输出一段连续文本，段落间用===隔开，不含换行符；请勿输出任何坐标或点标记、bbox 等标注；严禁在段落前重复“空间核心骨架”“地面肌理”等标题；严禁使用“1.”“2.”等序号

为验证视觉语言的准确性，结合 Prometheus-Vision 模型[18]和人工审核对随机抽取的 50 个样本进行准确性验证，模型和人工评估在同一标准下（表 2）展开。模型评分平均值为 4.12，人工评分平均值为 4.26，未出现严重偏离事实样本，生成文本仅在细节描述的详细情况上有差异，说明在提示词引导下的视觉大语言模型生成的描述具有较高的准确性与可信度。

表 2 评估标准

分值/分	要求
1	描述完全未能识别主要的居住区组件,或完全曲解了整个场景(如将其误认为是现代商业区或公园);对建筑环境存在重大的事实性错误
2	描述仅识别出一两个基本要素(如“楼房”“道路”),但遗漏了老旧小区典型的显著元素;缺乏任何关于材质、纹理或老化迹象的细节
3	描述正确识别社区场景的主要要素(如住宅楼、铺装道路、植被);但描述非常笼统,忽略关于风化、材质状况或具体生活痕迹的细节(如仅描述“旧墙”,未描述“带有剥落油漆的污渍混凝土墙”)
4	描述准确识别大多数关键要素,并提供关于其材质、纹理或场景空间特征的一些细节;它注意到老化或杂乱的迹象(如外露电线、防盗窗),但可能遗漏一些次要的生活气息元素,或描述不够丰富
5	描述精确且全面地识别所有关键要素,描述它们的形式、材质和纹理(如“斑驳的砖墙立面”“纠缠的架空电线”),并捕捉到整体的空间特征(如“狭窄封闭的巷道,充满浓厚的生活气息”)

2.2.3 BERtopic 主题建模

采用 BERTopic 框架对老旧社区图像描述文本的 4 个维度分别进行聚类。BERTopic 是一个模块化的神经主题建模框架，通过嵌入、UMAP 降维、HDBSCAN 聚类、在表示层利用 c-TF-IDF 算法等一系列步骤构建语义连贯的空间特征簇[19]；使用 Bge-M3 模型对文本进行嵌入，并在表示层增加 BM25 加权机制与高频词自动降权算法优化冗余词汇权重；选取 C_v（语义一致性）、C_{npmi}（归一化点互信息）、U_{mass} 以及主题多样性（Topic Diversity）4 个指标对聚类结果进行评估[20-22]，通过网格搜索法，对比不同聚类策略下邻居数与最小样本数量在 10~40 的参数组合，找到平衡数学统计的一致性与人类感官可读性的最优参数，最终获得各维度下的最优聚类方案（表 3）。4 个维度感知主题的 C_v 分值均处于 0.52~0.61，反映出聚类结果具有极高的语义连贯性与人类可解释性；C_{npmi} 指标均为正值，表明主题内部词项具有合理的统计相关性与逻辑一致性；U_{mass} 维持在稳健的负值区间，主题多样性均在 0.5 以上，表明生成了丰富且语义相对独立的主题。

表 3 参数选择与指标

维度	配置参数	主题数量	噪音数量	指标
空间	邻居数:15 最小样本数量:20 主题选择策略:eom 表示模型:MMR,diversity=0.4	16	398 (24.8%)	C_v=0.52 C_{umass}=-2.00 C_{nmpi}=0.05 多样性=0.58
细节	邻居数:20 最小样本数量:35 主题选择策略:leaf 表示模型:MMR,diversity=0.3 启用主题种子词:步道 小径 石阶 地砖 铺装 沥青 公路 主路 车道 宽阔 街道 商铺 贩卖棚屋 经营 花园 花坛 盆栽 绿植 院落 石凳 石桌 木椅 长椅 座椅 健身器材 单杠 漫步机 锻炼 停车场 停车位 整齐 停满 汽车 行人 穿行 混杂 过路 外墙 立面 窗户 防盗网 阳台 垃圾桶 杂物 堆放 废弃 纸箱 配电箱 电线 管道 井盖 标识牌	12	644 (40.1%)	C_v=0.61 C_{umass}=-2.11 C_{nmpi}=0.08 多样性=0.62
绿化	邻居数:25 最小样本数量:25 主题选择策略:leaf 表示模型:MMR,diversity=0.5	14	708 (44.6%)	C_v=0.52 C_{umass}=-2.51 C_{nmpi}=0.06 多样性=0.69
氛围	邻居数:20 最小样本数量:40 主题选择策略:leaf 表示模型:MMR,diversity=0.4	12	611 (38.5%)	C_v=0.56 C_{umass}=-2.49 C_{nmpi}=0.04 多样性=0.71

由于 Bertopic 倾向于捕捉具有显著特征的文本描述，24.8%～44.6%的样本因文本的说明性不足被划分为噪音。为弥补单一文本模态对模糊语义处理的不足，通过模型的视觉能力对部分噪音样本进行找回。基于 Visualized-BGE 的跨模态语义对齐能力，将原始图像向量映射至文本聚类空间以重分类部分噪声样本，将噪音样本控制在 2.6%～11.9%；结合每张图像的地理信息在 ArcGIS 中以 10 m×10 m 栅格为基本单元在单个社区中进行空间热点探测，以揭示 4 个感知维度特征的空间分布格局。

2.2.4　特征空间识别

利用多元聚类 Multivariate clustering 方法处理 4 个维度下的 16 种特征要素，分析各要素在空间中的共现结构，识别老旧社区中由特定特征组成的功能区域。由于不同视觉类型出现频率差异巨大，选取热点分析生成的 Z-score（标准差倍数）作为聚类输入变量。Z-score 反映网格及其邻域的统计显著性，能够过滤掉离群分布的视觉噪声，并将所有维度标准化为同一尺度的参数进行计算。

3　结果与分析

3.1　文本聚类和多模态找回

3.1.1　特征分析与分类

排除噪音聚类后，4 个维度聚类共计生成 50 个主题；综合文档分布图、主题距离图、层次聚类图，以及主题特征文档，将分散的主题整合为 17 个小组，对重庆老旧社区的特征进行描述（图 1）。

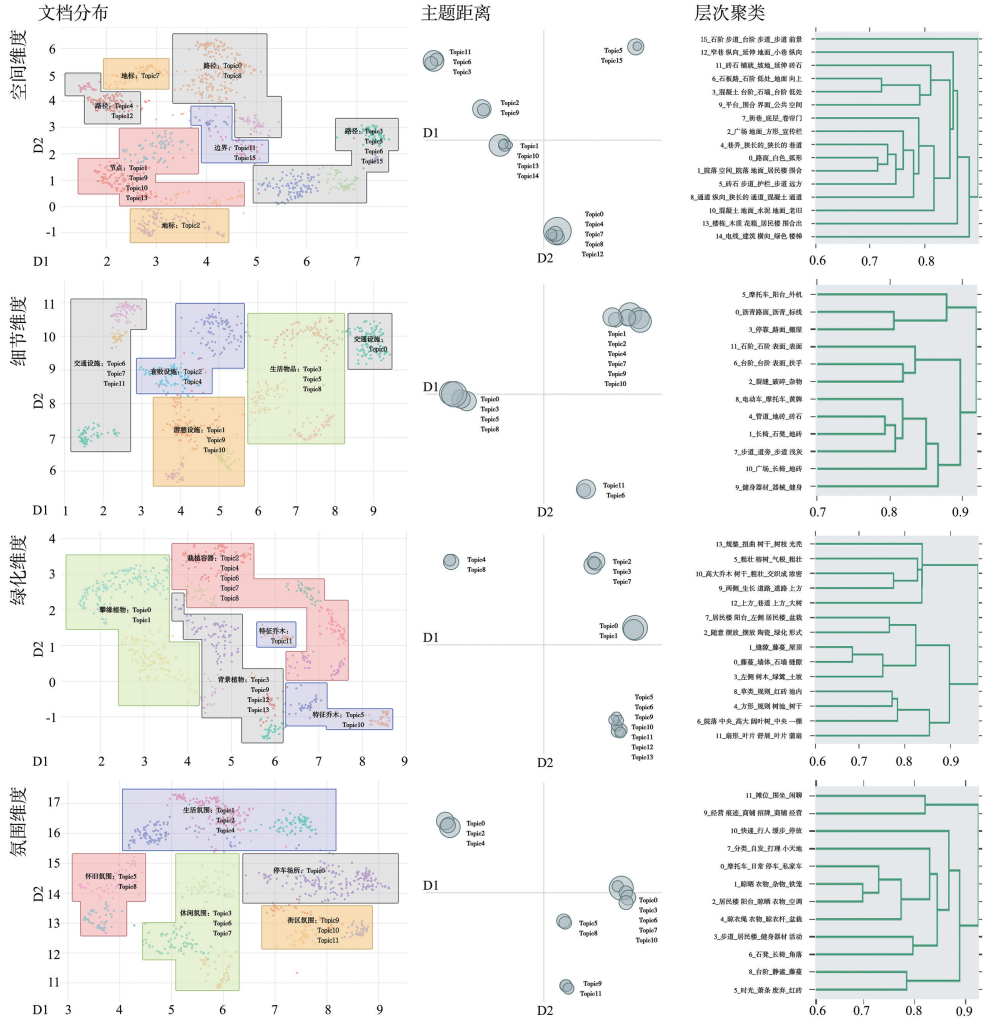


图 1　聚类结果可视化

1）空间维度生成 15 个主题，总结为 4 组。地标：包含指向广场、运动场（Topic2）和商业街巷（Topic7）的两个空间主题，这些场所的功能属性突出，记忆点强，是老旧社区居民的重要

活动目标点，起到方向定位的作用。节点：老旧社区中一些人群聚集点，包含院落（Topic1、Topic13）、平台或水泥坝（Topic10），这些场所不如商业或者运动场地功能突出，设施较少，但具有提供停留、聚集、交流、社交的功能。路径：居民在老旧社区中移动的线性空间，包括硬质化的道路（Topic0、Topic5、Topic6）、竖向连接阶梯（Topic3、Topic15）、狭窄的巷道（Topic4、Topic8、Topic12）。边界：分隔空间的硬质界面，包含社区边缘的坡体、围挡（Topic11）和内部居民楼建筑（Topic14）。

2）细节维度生成 11 个主题，总结为 4 组。交通设施：物品较少，以道路为主的空间，包含沥青道路（Topic0）、地砖步道（Topic7），较完整或较陈旧的台阶（Topic6、Topic11），其中道路的材质特征在该组中较为突出。活动设施：以活动和游憩设施为主的空间，包含户外家具（Topic1）、健身设施（Topic9）、广场空间（Topic10）。生活物品：包含大量物品描述的空间，其中两轮电动车、三轮车成为突出的关键词（Topic3、Topic5、Topic8），同时伴随垃圾回收站、衣物等生活杂物，是距离生活空间较近、堆放私人物品或停放小型载具的场所。衰败设施：以破损、老旧的墙体、铺装为突出要素，包括堆积杂物（Topic2）、裸露管道（Topic4）的空间。

3）绿化维度生成 13 个主题，总结为 4 组。攀援植物：指依附于墙面或墙体破损形成的裂缝生长的攀援植物或蕨类植物（Topic0、Topic1），反映一定的自生植物特征。特征乔木：以某一类型的乔木作为最突出的绿化特征，可能是该场景的主景树或主要树种，包含榕树（Topic5）、大乔木（Topic10）、棕榈植物（Topic11）。栽植容器：基于人工容器进行栽植的植物景观，如居民通过盆栽布置的阳台或花园（Topic2、Topic7），或是以种植池形式进行绿化的院落景观（Topic4、Topic6、Topic8）。背景植物：由植物形成界面，如密集植物形成的背景或绿篱（Topic3、Topic13），沿街的行道树或巷道两旁的树木（Topic9、Topic12）。

4）氛围维度生成 11 个主题，总结为 5 组。怀旧氛围：风化陈旧、被植物覆盖的砖制结构（Topic5）或石制台阶（Topic8），呈现陈旧但统一的老城质感。生活氛围：居民楼向外展示的生活空间（Topic1、Topic2、Topic4），各种衣物、盆栽、物品呈现出浓郁的生活气息。休闲氛围：生活区附近的休闲空间，如运动场所（Topic3）、可以停留的社交空间（Topic6），以及居民自发打理的庭院空间（Topic7）。街区氛围：生活区附近人流较大的场所，呈现一种热闹氛围，包括商铺（Topic9）、摊贩（Topic11）、行人较多的场所（Topic10）。停车场所：被两轮电动车或轿车停放占据的空间（Topic0）。

3.1.2 老旧社区特征的空间分布

以电力小区为例，不同维度特征要素的热点呈现不同分布状态（图 2）。空间维度：节点、地标相互之间呈现明显边界，被路径进行连接；路径和地标存在部分重合，呈现该社区中的重要空间直连主要道路的特点。细节维度：衰败设施和交通设施、游憩设施在空间上存在重合，反映老旧社区中通行空间视觉质量欠佳、部分游憩设施陈旧的特点。绿化维度：特征乔木和栽植容器在空间中较为集中，反映社区中以某种植物为主的绿化空间和居民搭建的绿化区域；攀援植物、背景植物则显示出线性分布特点，可能依附于道路或建筑墙体。氛围维度：社区的西南区域氛围呈现出明显的融合特征，反映该区域可能承载丰富的活动；东部区域则生活和休闲氛围更浓郁，反映该区域更偏向于居民日常生活使用。

3.2 特征空间共现区域可视化

3.2.1 特征空间共现区域可视化

基于伪 F 统计量、轮廓系数与间隔统计量的评估结果，本研究将研究区域划分为 8 类（Cluster=8）典型的空间共现模式（图 3）。通过变量诊断分析可知，怀旧氛围（$R^2=0.732$）、休闲氛围（$R^2=0.646$）、攀援植物（$R^2=0.645$）是主导本次空间聚类的核心因子。这表明老旧社区的空间异质性首先体现为历史风貌保留程度的差异，其次是公共生活活力的分异；而单纯的物理设施（如交通、停车）对空间类型的解释力相对较弱。排除每个维度值都低于平均值的背景区域聚类 5，基于多元箱型图（图 4~10）对 7 个聚类区域进行分析，Bertopic 聚类主题到空间共现聚类区域的映射过程如图 11 所示。

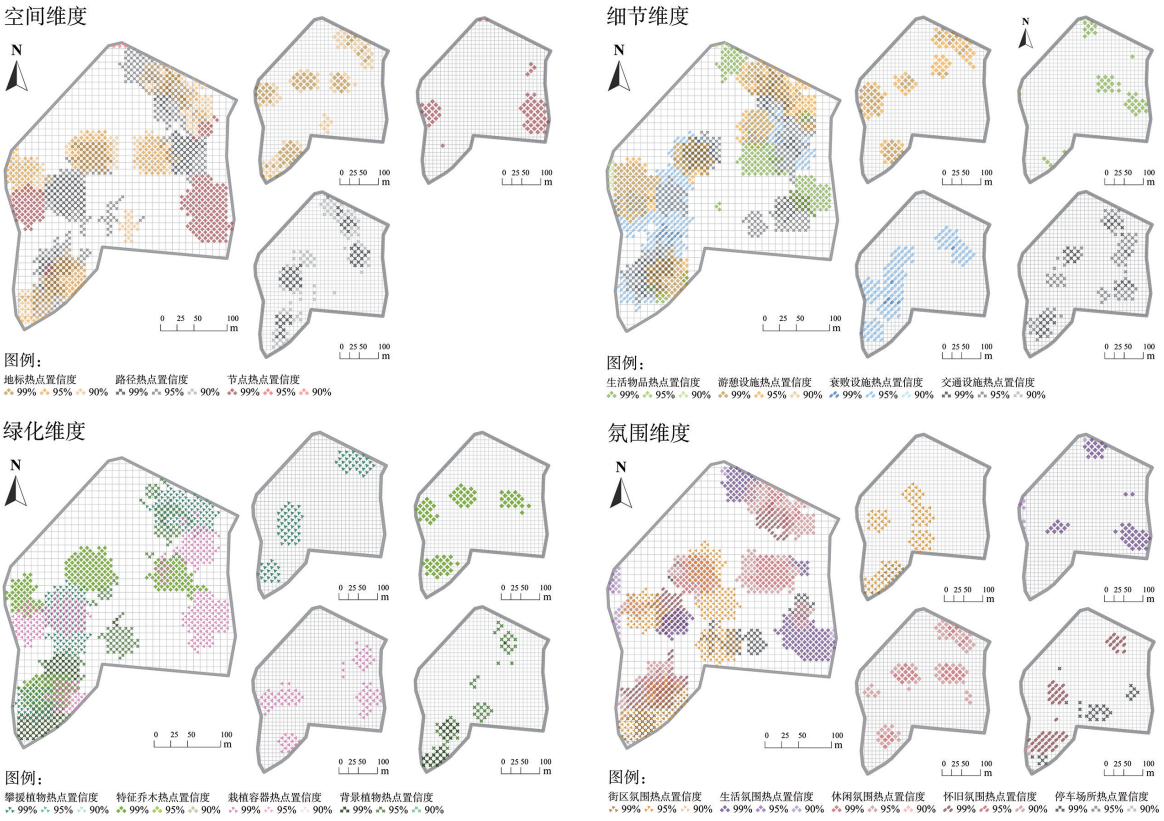


图 2　老旧社区特征的空间分布可视化

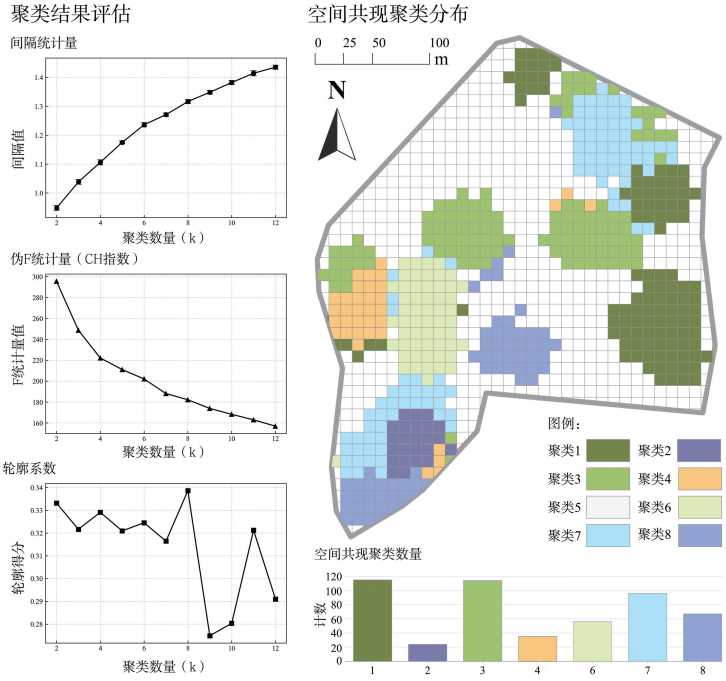


图 3　老旧社区特征空间共现区域可视化

1）空间共现聚类 1：邻里生活区域。该聚类表现为具有良好维护的生活场景。其中，节点、生活氛围、栽植容器的 Z-score 显著高于全域平均水平，反映该场景可能是一个靠近居民楼，被充分利用的院落或者花园空间，较高的生活物品 Z-score 值表明该区域还存在电动两轮车停放、垃圾回收站、告示牌等生活服务设施。

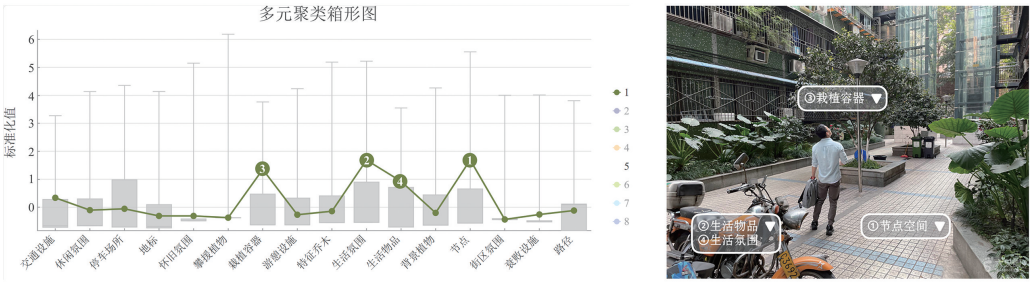


图 4 邻里生活区域–多元聚类箱形及区域代表图像

2）空间共现聚类 2：怀旧漫步街道。该聚类表现为具有游憩、一定商业活动、开放的老旧街区。其中，街区氛围、怀旧氛围、游憩设施、交通设施的 Z-score 值较高，反映其立面陈旧的街道空间特征；较高的背景植物则描述其具备组成界面或依附界面的绿化特点。

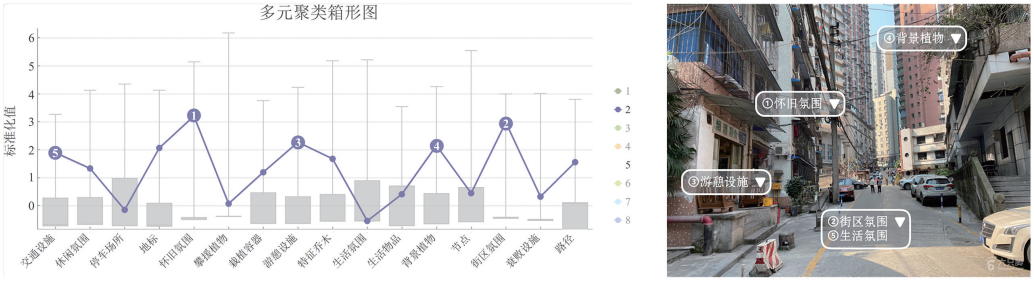


图 5 怀旧漫步街道–多元聚类箱形及区域代表图像

3）空间共现聚类 3：休憩游园区域。该聚类表现为自然环境良好的休闲场所。其中，休闲氛围、特征乔木、地标的 Z-score 值较高，反映其以有识别性的植物空间和休闲活动功能为核心特征；较高的游憩设施 Z-score 说明其不是单纯的绿化空间，而是为居民提供荫蔽的活动空间。

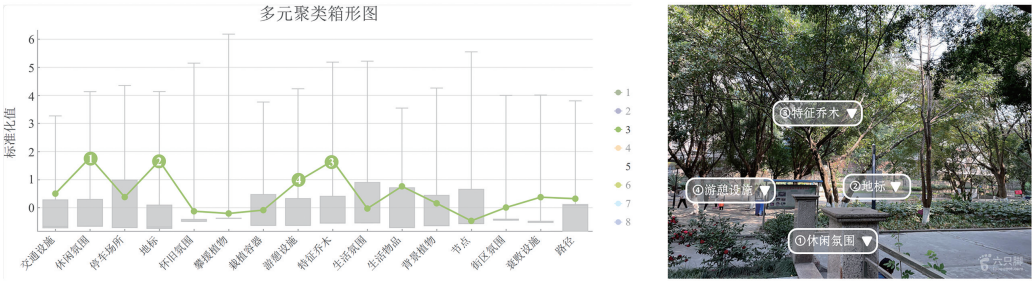


图 6 休憩游园区域–多元聚类箱形及区域代表图像

4）空间共现聚类 4：社区活力广场。该聚类表现为公共游憩活动场所。其中，游憩设施的 Z-score 值最高，表明该场所存在桌椅或健身器材等游憩设施；突出的节点、栽植容器 Z-score 值则说明该场景作为公共空间维护良好；较高的街区氛围 Z-score 值则说明该空间为居民提供了一个社交场景，成为社区中的活力点。

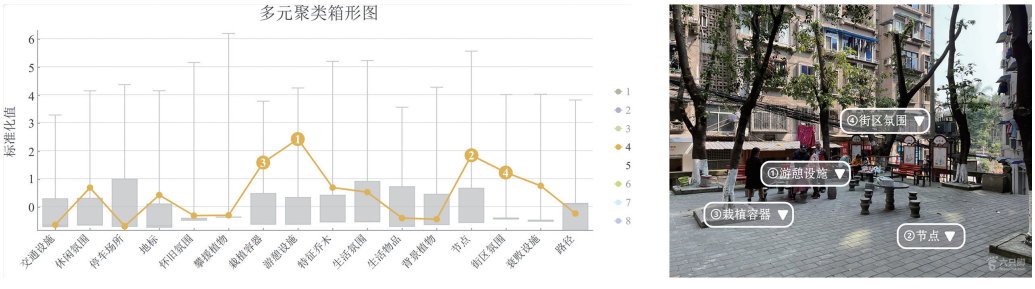


图 7 社区活力广场–多元聚类箱形及区域代表图像

5）空间共现聚类 6：老旧立面巷道。该聚类表现为具有建筑边界的线性空间。其中，攀援植物、路径、怀旧氛围的 Z-score 值突出，反映其视觉特征由附着攀援植物的建筑立面主导，形成了极具辨识度的具有年代感的巷弄景观；同时，显著的衰败设施 Z-score 值说明空间内伴随着陈旧的墙体、生活杂物或外露管线，这是社区中具有突出老旧质感的步行场景。

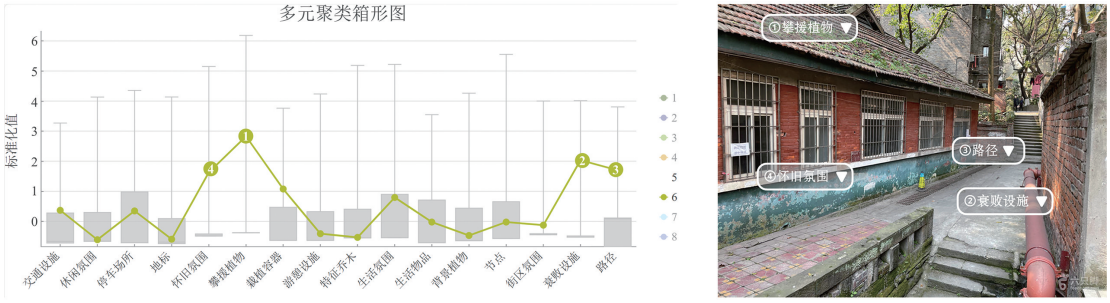


图 8　老旧立面巷道–多元聚类箱形及区域代表图像

6）空间共现聚类 7：绿化步行通廊。该聚类表现为被植物包围的线性空间。其中，较高的路径、交通设施、怀旧氛围 Z-score 值反映其同样是老旧社区内线性交通骨架；但较高的攀援植物、背景植物 Z-score 值反映了该聚类具备更明显的绿化特点，步道的一侧或两侧被植物覆盖、包裹，成为更具生态感的绿色步道。

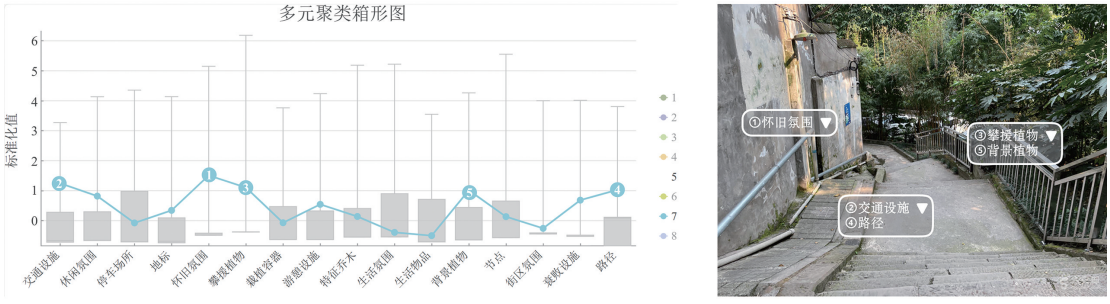


图 9　绿化步行通廊–多元聚类箱形及区域代表图像

7）空间共现聚类 8：边缘停车区域。该聚类表现为老旧社区边缘具有行道树的外部街道。其中，较高的停车场所、街区氛围、背景植物 Z-score 值共同说明了该街道上沿街的绿化特点和停车的行为特征。

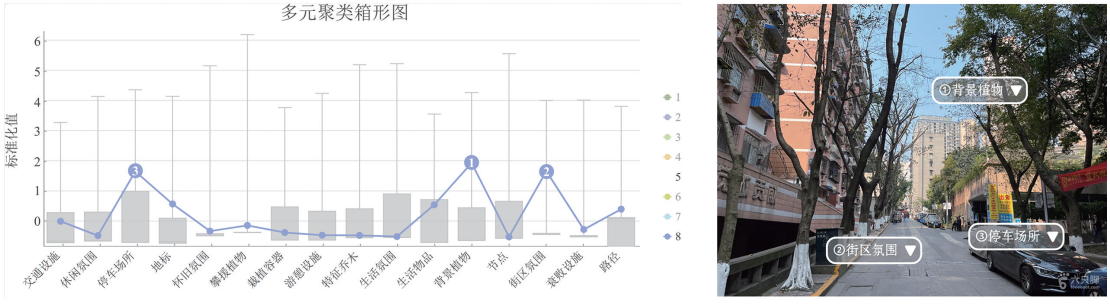


图 10　边缘停车区域–多元聚类箱形及区域代表图像

3.2.2　空间聚类改造建议

1）基于空间特征分布的结构整合。通过 7 类空间区域的地理分布，识别出该社区以中心休憩游园区为核心，连接西南、东北两条路径向外延展的整体空间特征。西侧由特色巷道串联街区和游园、广场等主要节点，宜考虑整体更新为社区文化艺术门户；东侧以邻里生活区域和自然休闲空间为主，宜进行针灸式更新打造邻里花园。

2）基于特征强度识别的要素补足。根据不同

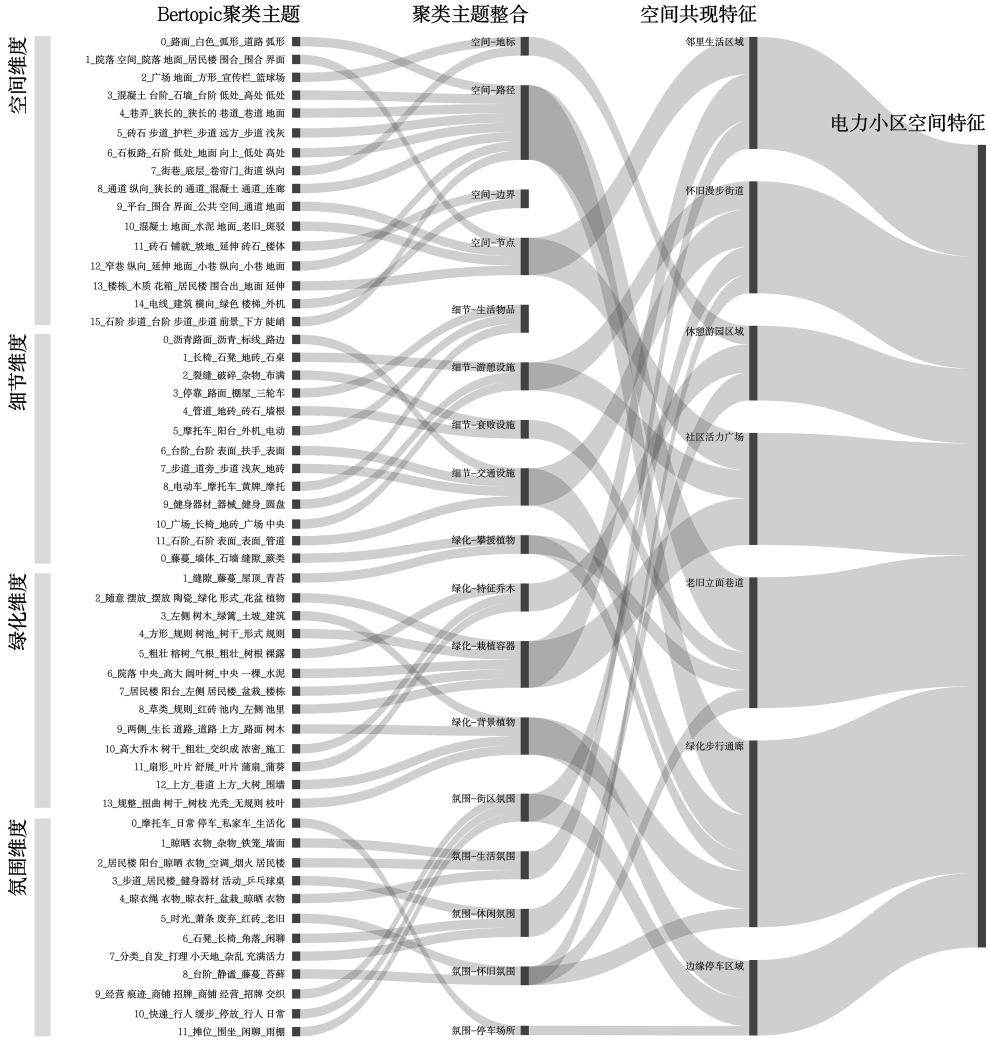


图 11 聚类主题到聚类区域映射过程

特征的 Z-score 值，可以判断空间薄弱点，明确改造方向，如邻里生活区域生活气息浓厚，但游憩设施不足，宜考虑适当规整生活杂物，结合现有景观要素进行补充。

3）基于特征组合优化的质量提升。特征的组合关系能够说明空间质量，助于深化改造，如老旧立面巷道的衰败设施特征与高怀旧氛围反映其风化材质的独特岁月感，应加以保留，避免覆盖式更新；而社区活力广场不具备明显的怀旧氛围，说明其衰败设施主要是物理破损，需要进行置换与美化。

综上，基于聚类特征的空间共现区域进行分类引导，为整体特征识别到针对性策略梳理提供依据，有助于挖掘每个社区自身的空间类型现状的问题点，实现具体空间具体分析，不同社区采用不同策略的更新方法。

4 结论与展望

本研究构建了基于多模态视觉语言模型与 BERTopic 聚类的老旧社区环境特征识别框架，利用众源图像对重庆市中心城区老旧社区进行语义特征空间共现分析，得出以下主要结论：

1）重庆老旧社区并非均质的衰败空间，而是具有异质性的复合空间。7 类典型的空间共现模式反映其核心差异非单纯由物理设施区分，而是源于历史风貌与公共生活的深度融合的“人本”特征。通过识别差异化共现组合中各类特征的多维价值，有助于明确微更新改造的目标，充分保留融入情感价值及社区记忆的存量要素。

2）多模态视觉语言模型能够高效的精细量化老旧社区环境特征。研究表明以 Doubao-Seed 为代表的 VLM 模型在多模态数据上训练出的图像语

义分析能力能够迁移到老旧社区的环境认知中，结合 BERTopic 与多元空间聚类技术精确捕捉特征在空间中的分布。该框架能够低成本、高通量利用网络媒体信息分析城市空间特征，有效缓解了单模态分析在老旧社区中社交媒体文本较少、图像视觉特征聚类受主观解释影响的局限，提高空间品质与居民体验等软维度评价的精细度，为老旧社区差异化更新提供更清晰、高效的技术支撑。

3）本研究仍存在一定局限性。受限于众源数据的空间分布不均与关注对象的非均质性，聚类结果可能存在偏差；且基于静态图像对现状特征的识别与描述难以捕捉时间动态变化与跨地域文化差异。未来研究将进一步融合居民偏好等多源数据，构建从“特征识别”到“潜力评价”再到“更新导则”的完整闭环，真正服务于“好房子、好小区、好社区、好城区”的系统性建设。

参考文献

[1]李有芳,李天翼,吴冠萌,等.老旧小区户外活动空间活力评价与更新策略研究[J].建筑学报,2023(增刊 2):174-180.

[2]施生旭,官烨苗.未来社区理念下老旧社区改造提升及评价指标体系构建[J].城市发展研究,2024,31(6):134-140.

[3]罗丹,罗融融.三峡库区移民社区公共空间景观偏好:以重庆万州周家坝街道为例[J].中国城市林业,2021,19(3):30-36.

[4]LIANG Y S, QIAN Q K, LI B, et al. A critical assessment on China's old neighborhood renovation: barriers analysis, solutions and future research prospects[J]. Energy and Buildings, 2025, 332:115407.

[5]骆骏杭,黄瓴.时空辩证与价值重塑:关于我国城镇老旧小区改造的冷思考[J].城市发展研究,2022,29(8):84-92.

[6]张春彦,张新翊,侯天颖.天津旧城更新中日常景观的记忆挖掘及空间转化[J].风景园林,2025,32(3):23-31.

[7]孟春霞,何捷.自发地理信息(VGI)数据中的宁海国家登山健身步道徒步者景观资源偏好[J].风景园林,2020,27(8):103-108.

[8]XIAO Y, TANG Y W. Can ChatGPT-4o assess the perceptions of streetscape change evidence from Shanghai, China[J]. Sustainable Cities and Society, 2025, 130:106674.

[9]LIN X H, YANG T, LAW S. From points to patterns: an explorative POI network study on urban functional distribution[J]. Computers, Environment and Urban Systems, 2025, 117:102246.

[10]赵烨桦,吴晓华.基于点评数据的城市公园社会服务评价[J].中国城市林业,2021,19(1):133-136.

[11]范雅婷,李越凡,涂伟,等.社交媒体数据支持的城市承灾体脆弱性评估:以深圳极端天气为例[J].城市规划,2024,48(8):101-113.

[12]CHAI-ALLAH A, FOX N, Günther F, et al. Mining crowdsourced text to capture hikers' perceptions associated with landscape features and outdoor physical activities[J]. Ecological Informatics, 2023, 78:102332.

[13]MA H R, LI J, Ye X Y. Deep learning meets urban design: assessing streetscape aesthetic and design quality through AI and cluster analysis[J]. Cities, 2025, 162:105939.

[14]GUO S, JANG K M, Duarte F, et al. Urban visual uniqueness: a landmark-free framework to quantify city's identity and distinctiveness from everyday scenes[J]. Computers, Environment and Urban Systems, 2025, 122:102351.

[15]ZHU H X, CHANG J, AN X Y, et al. Global and local feature extraction of urban historical spatial perception using large language models: a case study of Harbin Central Street District[J]. Cities, 2025, 165:106183.

[16]ZHANG X, CHEN M Z, HUANG Y M. Who gets to use the street Evaluate the utilization and inclusiveness using crowdsourced videos and vision-language models[J]. Sustainable Cities and Society, 2025, 134:106906.

[17]TUNG Y H, YANG Z R, SHEN M W, et al. Research note: assessing human preferences for natural landscapes: an analysis of ChatGPT-4 and LLaVA models[J]. Landscape and Urban Planning, 2025, 259:105371.

[18]LEE S, KIM S, PARK S. Prometheus-Vision: Vision-Language Model as a Judge for Fine-Grained Evaluation[C]//LUN W K, MARTINS A, SRIKUMAR V. Findings of the association for computational linguistics: ACL 2024. Bangkok, Thailand: Association for Computational Linguistics, 2024.

[19]GROOTENDORST M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure[EB/OL]. (2022-03-11)[2026-01-01]. https://arxiv.org/abs/2203.05794.

[20]RÖDER M, BOTH A, HINNEBURG A. Exploring the space of topic coherence measures[EB/OL]. (2015-02-02)[2026-01-01]. https://www.tesble.com/10.1145/2684822.2685324.

[21]MIMNO D, WALLACH H, TALLEY E, et al. Optimizing semantic coherence in topic models[C]//Proceedings of the 2011 conference on empirical methods in natural language processing. Association for computational linguistics, 2011.

[22]DIENG A B, RUIZ F J R, BLEI D M. Topic modeling in embedding spaces[J]. Transactions of the Association for Computational Linguistics, 2020, 8:439-453.